Educational
Research Advances
Vol.1, No.2, p. 96-106, 2026

Original article

Cultivating Evidence-Based Reasoning Through Textbook Design: A Comparative Analysis of Chinese and English Science Textbooks Based on the Toulmin's Argument Pattern

Chaoyue Zhuang, Cunkuan Wang*

College of Teacher Education, Ningbo University, Ningbo 315211, P. R. China

Received: May 13, 2026
Revised: June 5, 2026
Accepted: June 18, 2026
Published: June 23, 2026

Cited as:
Zhang, C., & Wang, C. (2026). Cultivating Evidence-Based Reasoning Through Textbook Design: A Comparative Analysis of Chinese and English Science Textbooks Based on the Toulmin's Argument Pattern. *Educational Research Advances*, *1*(2), 96-106.
https://doi.org/10.46690/era.2026.02.03

Abstract: Evidence-based reasoning (EBR) is a core competency in science education, yet little is known about how textbooks—the primary instructional resource—operationalize this competency. This study develops a three-dimensional analytical framework based on the Toulmin' s Argument Pattern to systematically compare the presentation of EBR in the topic "microscopic structure of matter" between the Zhejiang Education Edition (China) and Cambridge Lower Secondary Science (international) textbooks. Using descriptive statistics, Mann-Whitney U tests, Spearman correlation, and ordinal logistic regression, the findings reveal that the two textbooks exhibit high overall consistency in EBR design, with the Cambridge edition showing marginal advantages in diversity of reasoning types, explicitness of warrants, and design of reflection prompts. The three elements—evidence, reasoning, and claims—demonstrate strong coherence, with completeness of reasoning processes identified as the only significant predictor of claim precision. These findings offer empirical insights for competency-oriented textbook development.

Keywords: Evidence-based Reasoning; Toulmin's Argument Pattern; Science Textbooks; Comparative Analysis

1. Introduction

In an era of globalization and information abundance, fostering scientific literacy has become a central goal of education systems worldwide. The Organisation for Economic Co-operation and Development (OECD), in its PISA 2025 Science Framework, explicitly articulates the need for future citizens to possess the ability to "critically interpret scientific data and evidence" (OECD, 2023), underscoring the critical role of evidence-based reasoning (EBR) in navigating complex information landscapes. Similarly, the U.S. Next Generation Science Standards (NGSS) identifies "engaging in argument from evidence" as one of eight essential science and engineering practices, with its assessment framework conceptualizing evaluation as a process of reasoning from evidence (Duncan et al., 2018). A cross-national analysis of curricula across 27 countries reveals that EBR has become a shared concern in science education, with curricula consistently emphasizing systematic processes of "reasoning based on evidence" (Lu, 2011).

In China, the Compulsory Education Science Curriculum Standards (2022 Edition, Revised 2025) identify "scientific thinking" as a core competency, within which "argumentation and reasoning" requires students to "establish relationships between evidence and explanations based on evidence and logic" (Ministry of Education of the People's Republic of China, 2020). This framing positions EBR not merely as a disciplinary method but as a foundational thinking competency for adapting to future society and cultivating innovative talent.

Curriculum standards serve as primary guidelines for textbook development (Zhang, 2021). Students' initial construction of scientific knowledge and their imitation of scientific thinking patterns largely begin with textbooks. Consequently, how textbooks present scientific concepts, organize inquiry activities, and articulate the relationship between evidence and conclusions directly shapes students' opportunities to develop EBR competencies. Currently, Chinese junior secondary science textbooks are undergoing competency-oriented revision. In September 2024, the Zhejiang Education Edition of the Compulsory Education Textbook: Science for Grade 7, compiled according to the new curriculum standards, was officially introduced. A distinctive feature of this new edition is its "core competency as

*Correspondence author.
E-mail: 1914274988@qq.com (C. Zhang); wcklx@nbu.edu.cn (C. Wang)

the main thread” design philosophy (Li & Lu, 2024). However, systematic, operational frameworks for analyzing how textbooks integrate EBR into content organization, activity design, and case presentation remain limited. As these new textbooks enter classrooms, teachers face uncertainties regarding their use, making systematic investigation of these textbooks highly relevant.

International science textbooks, such as the Cambridge Lower Secondary Science developed by Cambridge Assessment International Education, have long emphasized “Thinking and Working Scientifically” as a core pillar, accumulating substantial experience in organizing and structuring EBR. First published in 2012 and revised in its second edition in 2021, this series is used in over 10,000 schools across more than 160 countries. The experiences gained from its science education approach offer valuable lessons. By comparing these two representative textbooks, this study aims to reveal how different curriculum cultures operationalize EBR, providing reference points for Chinese textbook development and supporting the translation of curriculum ideals into textbook practice.

The topic “microscopic structure of matter” is a core chemistry theme in junior secondary science curricula, involving the construction of abstract concepts such as atoms, molecules, and elements. These concepts cannot be acquired through direct observation; students must reason from evidence—inferring the existence and movement patterns of microscopic particles from macroscopic phenomena, and deriving periodic patterns from experimental data such as atomic weights and valence states. This theme thus exemplifies the core value of EBR in science learning: evidence serves as a bridge to the invisible world, while reasoning functions as the cognitive tool for constructing abstract concepts.

Furthermore, this theme is systematically presented in both textbooks with comparable content structures: the Zhejiang Edition (Grade 7, Volume 2, Chapter 2 “Microscopic Structure of Matter”) and the Cambridge Edition (Stage 7, Unit 2.5 “Atoms, Elements and the Periodic Table,” and Unit 2.6 “Compounds and Chemical Formulas”) exhibit strong alignment in knowledge organization, providing a solid foundation for cross-version comparison.

2. Theoretical Framework

2.1 Defining Key Concepts: Evidence, Reasoning, and Argument

Before presenting the theoretical framework, it is necessary to clarify the three core concepts—evidence, reasoning, and argument—and their interrelationships.

Evidence refers to information that supports or challenges the truth of a proposition (Peng & Ma, 2010; Wang, 2019). In science education, evidence commonly appears as experimental data, observations, charts, case studies, and historical accounts of scientific discoveries. Evidence constitutes the raw material for reasoning, serving as the starting point of cognitive processes.

Reasoning is the cognitive process of deriving a proposition (conclusion) from one or more propositions (premises) (Lin, 1990). Reasoning manifests as logical connections among propositions. From an educational perspective, reasoning encompasses deductive, inductive, and analogical forms (Lin, 1990). In science education, reasoning involves not only following logical rules but also critically evaluating the sources and quality of evidence. It is precisely this “reasoning based on evidence” that distinguishes scientific reasoning from purely logical reasoning.

Argument is the structured presentation of reasoning, comprising a claim, evidence, and the logical links connecting them (Peng & Ma, 2010). In science education, arguments often take the form of explicit, logical expressions organizing evidence, warrants, and rebuttals around scientific claims. Argument is the external representation of reasoning, enabling the communication and evaluation of cognitive products.

In this study, these three concepts are treated as analytically distinct yet closely interrelated, as illustrated in Figure 1: evidence provides the foundational material, reasoning constitutes the cognitive activity linking evidence to claims, and argument represents the externalized, structured form of reasoning. The Toulmin model is well-suited as an analytical tool because it transforms this internal cognitive process into analyzable structural elements.

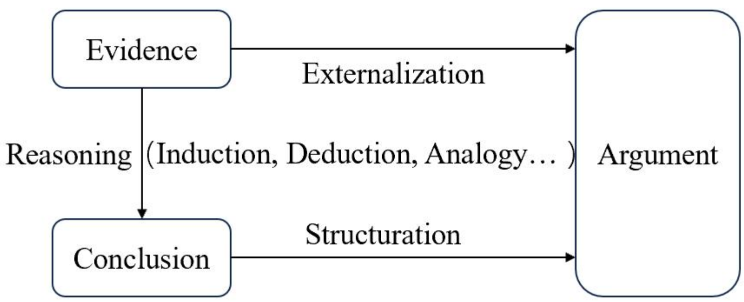

Figure 1. The interplay between evidence, reasoning, and Argument

2.2 The Toulmin’s Argument Pattern and Its Appropriateness for Analyzing Evidence-Based Reasoning

The Toulmin model of argumentation, introduced by British philosopher Stephen Toulmin in *The Uses of Argument* (1958), conceptualizes argument as a dynamic process akin to legal proceedings, moving beyond formal logic's syllogistic structures to better reflect real-world reasoning practices (Mo, 2004). The model comprises six core elements (Wu, 2003) .See Figure 2 for specific relationships: *Claim* (the conclusion to be justified), *Data* (facts or evidence serving as the argument's foundation), *Warrant* (the bridge connecting data to claim, explaining why the data support the claim), *Backing* (theoretical or authoritative support for the warrant), *Qualifier* (indicating the strength of the inferential leap, such as "possibly" or "under certain conditions"), and *Rebuttal* (acknowledging exceptions or competing explanations).

The Toulmin model's theoretical appropriateness for analyzing EBR lies in its direct mapping onto the cognitive process of reasoning from evidence. Data corresponds to evidence; claim corresponds to the conclusion; warrant corresponds to the logical rule linking evidence to conclusion; and backing, qualifier, and rebuttal correspond to considerations of credibility, conditionality, and criticality. Thus, the Toulmin model is fundamentally a tool for making the reasoning process explicit and structured, not limited to analyzing formal argumentation.

In textbooks, EBR often remains implicit, embedded within narrative descriptions, activity designs, or visual representations. The Toulmin model offers an operational framework for making these implicit reasoning chains explicit, enabling systematic identification, comparison, and evaluation. By analyzing the presence and completeness of Toulmin elements in textbooks, we can infer the extent to which textbooks support students' development of evidence-based reasoning.

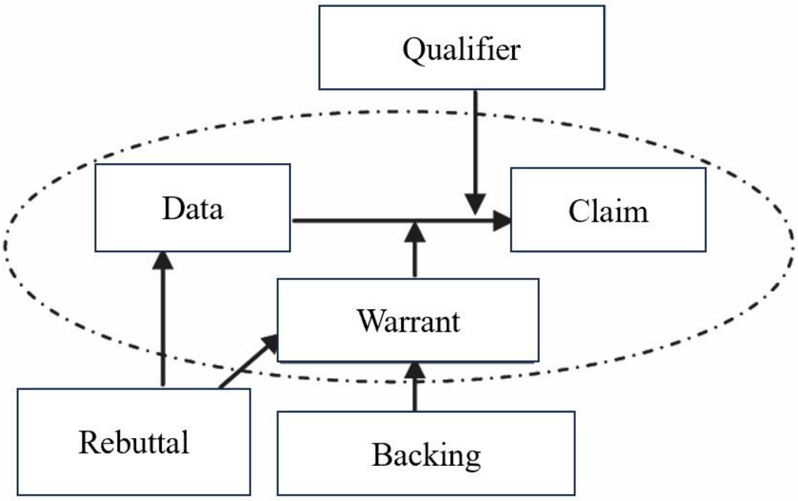

Figure 2. Toulmin's Argument Pattern (Toulmin, 1958)

2.3 Operationalizing Evidence-Based Reasoning: A Three-Dimensional, Six-Indicator Analytical Framework

To apply the Toulmin model to the analysis of junior secondary science textbooks, this study adapts and operationalizes the model based on the approach of Xu and Zhang (2023) and the characteristics of junior secondary science textbooks. The framework conceptualizes EBR in textbooks along three dimensions: sources and relations of evidence, processes and types of reasoning, and expression and reflection of claims. Table 1 presents the complete analytical framework.

Table 1. Analytical Framework for Evidence-Based Reasoning in Science Textbooks

Primary Dimension	Secondary Dimension	Levels
A. Multi-source and interconnectedness of evidence	A1. Diversity of evidence sources	1. No evidence source 2. Single evidence source 3. Dual evidence sources 4. Multiple evidence sources
	A2. Closeness of evidence relations	1. No evidence 2. Isolated evidence 3. Parallel evidence 4. Progressive evidence
B. Multi-step richness of reasoning	B1. Completeness of reasoning processes	1. Claim only (C) 2. Claim with evidence support (C+D) 3. Claim with evidence and warrant (C+D+W)

		4. Claim with evidence, warrant, and backing (C+D+W+B)
	B2. Diversity of reasoning types	1. No reasoning 2. Single reasoning type 3. Combination of two reasoning types 4. Comprehensive application of multiple reasoning types
C. Rigor and criticality of claims	C1. Precision of claim expression	1. No qualifier 2. Vague qualifier 3. Specific qualifier 4. Universal qualifier
	C2. Depth of claim reflection	1. No reflection prompt 2. Comprehension-focused reflection 3. Application-focused reflection 4. Critical reflection

The framework operationalizes each dimension as follows:

Dimension A – Multi-source and interconnectedness of evidence focuses on the quality of evidence provided to support scientific claims, examining both the diversity of evidence sources (categorized as everyday examples, experimental data, and historical cases) and the logical relationships among multiple evidence items (ranging from isolated to progressive).

Dimension B – Multi-step richness of reasoning examines how textbooks reveal or guide the logical connections between evidence and claims. Completeness of reasoning processes is assessed by the presence of Toulmin elements (C, D, W, B), while diversity of reasoning types captures the use of causal reasoning (identifying cause-effect relationships), analogical reasoning (using familiar situations to understand abstract concepts), and interpretive reasoning (making meaning from phenomena or data).

Dimension C – Rigor and criticality of claims focuses on how textbooks present scientific conclusions and cultivate students' reflective capacities. Precision of claim expression is assessed by the presence and specificity of qualifiers (e.g., "always," "under certain conditions"), while depth of claim reflection is assessed by the design of reflection prompts (ranging from comprehension to critical evaluation).

3. Research Design

3.1 Units of Analysis

The unit of analysis in this study is the "instructional meaning segment"—a collection of text, figures, and captions that together develop a specific scientific claim. An instructional meaning segment typically spans multiple paragraphs and constitutes a complete exposition or argumentation of a scientific conclusion. For example, when introducing "atomic structure," a textbook may first present experimental phenomena (evidence), then propose an atomic model (claim), and explain the connection between them (warrant). These components together constitute an instructional meaning segment.

The instructional meaning segment was selected as the unit of analysis for two primary reasons. First, the Toulmin model requires analysis of complete argumentation structures, necessitating identification of claims, evidence, and logical connections within a relatively independent context. Segmenting at the sentence level would fragment this structure, while analysis at the chapter level would obscure it. Second, instructional meaning segments reflect textbook authors' complete conceptualization of a knowledge point, preserving the original logical structure.

For the topic "microscopic structure of matter," analysis of the Zhejiang Edition (Grade 7, Volume 2, Chapter 2) yielded 30 instructional meaning segments, while analysis of the Cambridge Edition (Stage 7, Units 2.5 and 2.6) yielded 12 instructional meaning segments. These segments served as the foundation for coding all six indicators.

3.2 Data Analysis Methods and Tools

This study employs a mixed-methods approach combining quantitative and qualitative analysis.

Descriptive statistics were used to present overall characteristics of EBR presentation across the two textbooks, including distributions of evidence types, Toulmin elements, and reasoning types. These descriptive results served as the foundation for subsequent inferential analyses.

Mann-Whitney U tests were employed to compare differences between the two textbooks across the six dimensions. This non-parametric test was chosen over independent-samples t-tests because the data are ordinal (levels 1–4) and may violate normality assumptions.

Spearman rank correlation was used to examine relationships among dimensions. This non-parametric approach is appropriate for ordinal data, robust to non-normality, and aligns methodologically with the Mann-Whitney U tests.

Ordinal logistic regression was employed to identify factors predicting claim precision (C1) and reflection depth (C2). Ordinal logistic regression is appropriate for ordinal dependent variables. Separate models were constructed for C1 and C2, with A1, A2, B1, and B2 as predictors and textbook version as a control variable. Model fit (Pearson and deviance chi-square), pseudo-R^2 (Nagelkerke), and regression coefficients were reported.

Data analysis was conducted using NVivo 20 for qualitative coding and management, and IBM SPSS Statistics 26 for quantitative analyses.

4. Results

4.1 Descriptive Analysis of Evidence-Based Reasoning Presentation

Table 2. Distribution of Evidence Types

Evidence Type	Zhejiang Edition (n = 38)	Cambridge Edition (n = 19)
Everyday Examples	19 (50.0%)	13 (68.4%)
Experimental Data	9 (23.7%)	5 (26.3%)
Historical Cases	10 (26.3%)	1 (5.3%)

As shown in Table 2, the analysis of evidence source diversity (A1) is based on the distribution of evidence types across instructional meaning segments with identifiable evidence sources. For the Zhejiang Edition, 38 valid segments were identified, while the Cambridge Edition yielded 19 valid segments.

Evidence types were categorized into three categories: everyday examples, experimental data, and historical cases. The results indicate that in the Zhejiang Edition, everyday examples account for 19 instances (50.0%), experimental data for 9 instances (23.7%), and historical cases for 10 instances (26.3%). This relatively balanced distribution suggests that when introducing scientific conclusions, the Zhejiang Edition draws substantially from everyday experiences while also incorporating experimental evidence and historical context. In contrast, the Cambridge Edition relies heavily on everyday examples, which constitute 13 instances (68.4%), followed by experimental data with 5 instances (26.3%), and historical cases with only 1 instance (5.3%). This distribution reveals that the Cambridge Edition predominantly draws upon everyday experiences, with markedly lower diversity in evidence sources.

Table 3. Distribution of Toulmin Elements

Toulmin Element	Zhejiang Edition (n = 77)	Cambridge Edition (n = 33)
Claim (C)	30 (39.0%)	12 (36.4%)
Data (D)	25 (32.5%)	11 (33.3%)
Warrant (W)	19 (24.7%)	10 (30.3%)
Backing (B)	3 (3.9%)	0 (0.0%)

The completeness of reasoning processes(B1) was assessed based on the distribution of Toulmin elements across the two textbooks. The frequencies of four core elements—Claim (C), Data (D), Warrant (W), and Backing (B)—were counted for each textbook (see Table 3). For the Zhejiang Edition, a total of 77 reasoning elements were identified, comprising 30 Claims (39.0%), 25 Data (32.5%), 19 Warrants (24.7%), and only 3 instances of Backing (3.9%). For the Cambridge Edition, a total of 33 reasoning elements were identified, including 12 Claims (36.4%), 11 Data (33.3%), 10 Warrants (30.3%), and no instances of Backing (0.0%).

These data indicate that both textbooks place considerable emphasis on presenting claims and supporting data, yet pay insufficient attention to warrants—the explicit logical links connecting evidence to claims. Moreover, the presentation of backing is extremely limited in both textbooks. The Cambridge Edition shows a slightly higher proportion of warrants (30.3%) compared to the Zhejiang Edition (24.7%), suggesting a marginally greater effort to make the reasoning process explicit. However, it lacks any backing elements, whereas the Zhejiang Edition includes a minimal number of backing instances (3.9%), albeit at a very low proportion.

Table 4. Distribution of Reasoning Types

Reasoning Type	Zhejiang Edition (n = 23)	Cambridge Edition (n = 19)

Causal Reasoning	7 (30.4%)	2 (10.5%)
Analogical Reasoning	10 (43.5%)	10 (52.6%)
Interpretive Reasoning	6 (26.1%)	7 (36.8%)

The diversity of reasoning types (B2) was examined based on the reasoning approaches employed in the textbooks. Reasoning types were categorized into three categories: causal reasoning, analogical reasoning, and interpretive reasoning. The analysis was conducted on instructional meaning segments that contained explicit reasoning processes, yielding 23 valid segments for the Zhejiang Edition and 19 for the Cambridge Edition.

The results, presented in Table 4, show that in the Zhejiang Edition, causal reasoning appears in 7 instances (30.4%), analogical reasoning in 10 instances (43.5%), and interpretive reasoning in 6 instances (26.1%). Overall, analogical reasoning predominates, followed by causal and interpretive reasoning, with a relatively dispersed distribution across the three types. In the Cambridge Edition, causal reasoning appears in 2 instances (10.5%), analogical reasoning in 10 instances (52.6%), and interpretive reasoning in 7 instances (36.8%). This indicates a strong reliance on analogical reasoning, with interpretive reasoning as the secondary type and causal reasoning being notably scarce.

Both textbooks exhibit a preference for analogical reasoning, which facilitates students' comprehension of abstract concepts by connecting them to familiar contexts. However, excessive dependence on analogy may compromise logical rigor. The low proportion of causal reasoning suggests that both textbooks place limited emphasis on revealing the underlying mechanisms of scientific phenomena.

4.2 Comparative and Correlational Analyses

4.2.1 *Mann-Whitney U Test Results*

Table 5. Mann-Whitney U Test Results Across Dimensions

Dimension	Mean Rank (Zhejiang)	Mean Rank (Cambridge)	Mann-Whitney U	Asymptotic p-value (two-tailed)
A1	18.50	22.00	140.000	0.312
A2	18.42	22.17	137.500	0.268
B1	18.92	21.17	150.500	0.514
B2	19.27	27.08	112.000	0.051
C1	19.88	24.83	133.000	0.206
C2	19.69	23.50	140.500	0.328

To compare the overall differences between the two textbooks across the six dimensions, this study employed the independent-samples Mann-Whitney U test, which is suitable for ordinal data and non-normally distributed variables, as it determines significant differences by comparing the rank orders of the two groups. The test results are presented in Table 5.

In terms of mean ranks, the Cambridge Edition shows higher values across all six dimensions, suggesting a tendency toward a higher overall level of evidence-based reasoning (EBR) presentation. However, with the exception of B2 (diversity of reasoning types), the p-values for the remaining five dimensions exceed 0.05, indicating that, statistically, the differences between the two textbooks are not significant. This suggests that despite the Cambridge Edition's slight advantages in certain aspects, the two textbooks demonstrate high consistency in the overall design of EBR materials.

The p-value for B2 is 0.051, which is very close to the significance threshold of 0.05 and can be considered marginally significant. Combined with the mean ranks (Zhejiang Edition: 19.27; Cambridge Edition: 27.08), this indicates that the Cambridge Edition holds a slight advantage in the diversity of reasoning types. This finding implies that the Cambridge Edition tends to integrate multiple reasoning approaches when presenting scientific knowledge—such as combining analogical and interpretive reasoning—whereas the Zhejiang Edition is relatively more focused on single reasoning types. Greater diversity in reasoning types may help students understand scientific concepts from multiple perspectives and cultivate flexible scientific thinking. Although the difference does not reach strict statistical significance, this trend warrants attention and offers potential directions for improving textbook design in China.

4.2.2 *Spearman Correlation Analysis*

Table 6. Spearman Correlation Matrix

	A1	A2	B1	B2	C1	C2

A1	1.000	0.779**	0.761**	0.577**	0.567**	0.472**
A2	0.779**	1.000	0.615**	0.515**	0.520**	0.473**
B1	0.761**	0.615**	1.000	0.784**	0.632**	0.381*
B2	0.577**	0.515**	0.784**	1.000	0.396**	0.477**
C1	0.567**	0.520**	0.632**	0.396**	1.000	0.162
C2	0.472**	0.473**	0.381*	0.477**	0.162	1.000

Note: ** $p<0.01$, * $p<0.05$

To explore the underlying relationships among the dimensions, this study employed Spearman rank correlation analysis. Spearman correlation was selected over Pearson correlation primarily because the data are ordinal in nature and may not satisfy the assumption of normality. The correlation results are presented in Table 6.

The correlation matrix reveals three key patterns.

First, the three elements—evidence, reasoning, and claims—demonstrate high coherence. A1 and A2 show a strong positive correlation (r=0.779, p<0.01), indicating that greater diversity in evidence sources is associated with more closely organized relationships among evidence items, progressing from parallel to progressive structures. A1 and B1 exhibit a strong positive correlation (r=0.761, p<0.01), suggesting that multiple types of evidence contribute to the formation of more complete reasoning chains. B1 and B2 show a strong positive correlation (r=0.784, p<0.01), revealing an important pattern: complete Toulmin argumentation structures tend to require the integrated use of multiple reasoning types. B1 and C1 also show a strong positive correlation (r=0.632, p<0.01), indicating that conclusions derived through complete reasoning processes tend to have more clearly defined scopes and conditions of applicability.

Second, completeness of reasoning processes and diversity of reasoning types mutually reinforce each other. The strong correlation between B1 and B2 (r=0.784) suggests that when constructing complete reasoning chains, textbook authors should intentionally integrate causal reasoning, analogical reasoning, and interpretive reasoning to enhance the depth and breadth of argumentation.

Third, reflection activities exhibit relative independence. Notably, C2 shows no significant correlation with C1 (r=0.162, p>0.05), and the correlations between C2 and other dimensions are generally lower than those among other dimension pairs (ranging from 0.38 to 0.48). This finding indicates that the design of reflection activities depends more on other factors—such as textbook authors' instructional philosophies—rather than being determined solely by the quality of evidence and reasoning presentation. It also implies that even when the organization of evidence and reasoning for a particular knowledge point is less than ideal, textbooks can still guide students toward higher-order reflection through specially designed thinking and discussion prompts.

4.3 Ordinal Logistic Regression

4.3.1 *The integrity of the reasoning process is the only significant factor determining the rigor of the predicted conclusion.*

Table 7. Ordinal Logistic Regression Results for C1 (Precision of Claim Expression)

Variable	Coefficient (*B*)	*SE*	*Wald*	df	*p*	95% *CI*
A1	-.502	.894	.315	1	.574	[-2.253, 1.250]
A2	.596	.494	1.454	1	.228	[-.372, 1.564]
B1	2.802	.940	8.887	1	.003	[.960, 4.645]
B2	-.834	.535	2.427	1	.119	[-1.883, .215]
Textbook version (Zhejiang)	-.828	.763	1.178	1	.278	[-2.324, .667]

Note: Model fit: $\chi^2(5)$=28.897, p<0.001; Nagelkerke R^2=0.540; Parallel lines test p=0.287.

To investigate the key factors influencing the precision of claim expression (C1), this study conducted an ordinal logistic regression analysis with C1 as the dependent variable, A1, A2, B1, and B2 as independent variables, and textbook version as a control variable. The results show that only B1 (completeness of reasoning processes) reached statistical significance (p=0.003), with a positive coefficient (2.802), indicating that after controlling for other variables, B1 has a significant positive effect on C1. That is, the more complete the reasoning process, the more clearly the conditional qualifications of the claim are articulated (making it more likely to reach Level 3 or 4). The remaining variables (A1, A2, B2, and textbook version) were not significant (p>0.05). The regression coefficients and their significance are presented in Table 7.

Based on the data, three equations represent the cumulative probabilities for different levels of C1 (i.e., the probability of scores ≤ 1, ≤ 2, and ≤ 3):

When predicting the probability of C1 ≤ 1: $\text{logit}(P(C1 \le 1)) = 2.924 - 2.802 \times B1$

When predicting the probability of C1 ≤ 2: $\text{logit}(P(C1 \le 2)) = 4.964 - 2.802 \times B1$

When predicting the probability of C1 ≤ 3: $\text{logit}(P(C1 \le 3)) = 7.994 - 2.802 \times B1$

where B1 takes values from 1 to 4. The complete equation incorporating all independent variables (though non-significant) is:

$$\text{logit}(P(C1\leq j))=\theta_j-(-0.502\times A1+0.596\times A2+2.802\times B1-0.834\times B2-0.828\times I)$$
$$=\theta_j+0.502\times A1-0.596\times A2-2.802\times B1+0.834\times B2+0.828\times I$$

where I=1 for the Zhejiang Edition and 0 for the Cambridge Edition; $\theta_1=2.924$, $\theta_2=4.964$, $\theta_3=7.994$.

This finding reveals an important principle for textbook design: fostering students' precise and conditional understanding of scientific conclusions depends critically on presenting complete reasoning chains that allow students to experience the entire reasoning process, rather than relying solely on the quantity of evidence or diversity of reasoning types. This constitutes a crucial foundation for developing students' well-rounded and comprehensive capacity for evidence-based reasoning.

4.3.2 *The relative independence of the reflection stage*

An ordinal logistic regression analysis was conducted with C2 (depth of claim reflection) as the dependent variable, using the same set of independent variables and control variable. The overall model was significant ($\chi^2=15.85$, df=5, $p=0.007$), with a Nagelkerke R^2 of 0.336, indicating that the model explains approximately 33.6% of the variance in C2. However, none of the individual independent variables reached statistical significance ($p>0.05$). The p-value for B2 (diversity of reasoning types) was 0.079, approaching significance and suggesting a potential positive trend that warrants further investigation with a larger sample. Textbook version also showed no significant effect ($p=0.498$), consistent with the Mann-Whitney U test results.

These findings corroborate the Spearman correlation analysis, confirming that the design of reflection activities is relatively independent of the quality of evidence and reasoning. The absence of significant effects from A1, A2, B1, and B2 on C2 suggests that the depth of reflection is not automatically determined by the preceding reasoning process but rather results from intentional design choices by textbook authors. This carries important implications for textbook development: even when the presentation of evidence and reasoning for a particular topic is suboptimal, textbooks can still guide students toward higher-order reflection through well-designed thinking and discussion prompts. Conversely, neglecting the design of reflection activities may leave students at the level of passive acceptance, hindering the development of critical thinking, even when evidence and reasoning are presented comprehensively. The factors influencing reflection design are complex and may involve variables not included in the present model, such as textbook authors' pedagogical philosophies or the structure of instructional features. Future research could further explore these potential determinants.

5. Discussion

5.1 Main Findings

This study systematically compared the presentation of evidence-based reasoning (EBR) in Chinese and international science textbooks using a Toulmin-based analytical framework. Four main findings emerge.

First, the Cambridge Edition shows a marginal advantage in the diversity of reasoning types. The Mann-Whitney U test for B2 approached statistical significance ($p=0.051$), with the Cambridge Edition demonstrating greater use of multiple reasoning types in combination. This finding suggests that the Cambridge Edition more frequently integrates analogical, causal, and interpretive reasoning, potentially offering students more varied cognitive pathways for understanding scientific concepts. Research on the "evidence-based reasoning" competency in high school chemistry indicates that students' monolithic reasoning structures and suboptimal levels of reasoning based on evidence may be attributed to the fact that most experiments in textbooks are confirmatory in nature; as a result, students have insufficient opportunities to independently design experimental plans or contemplate what evidence is required to solve a problem, let alone to develop diverse reasoning abilities (Yan, 2019).

Second, the Cambridge Edition presents warrants more explicitly than the Zhejiang Edition. Descriptive analysis revealed that warrants constitute 30.3% of Toulmin elements in the Cambridge Edition compared to 24.7% in the Zhejiang Edition. While both textbooks underutilize warrants—the crucial "because" linking evidence to claims—the Cambridge Edition shows slightly greater attention to making these logical connections explicit. The reasonableness of this conclusion is further corroborated by other existing research findings. One study pointed out that the key component of the Toulmin argumentation model—the warrant—is often overlooked in practical teaching research, and its definition remains somewhat ambiguous (Liu, 2020), which may be a crucial reason why both textbook series lacked sufficient utilization of this argumentative element. Furthermore, another study compared the scientific argumentation elements in Chinese and American junior high school biology textbooks and found significant differences in the presentation of these elements between the two countries, with American textbooks demonstrating greater prominence in terms of argumentation completeness and logical coherence (Fang et al., 2023). This suggests that foreign textbooks may place more emphasis on the use of these argumentative elements compared to the domestic Zhejiang Edition.

Third, the Cambridge Edition designs reflection prompts more extensively. Although the difference in C2 did not reach statistical significance, qualitative analysis indicated that the Cambridge Edition's reflection questions more frequently prompt critical evaluation (e.g., "Do you think this conclusion holds in all circumstances?"), while the Zhejiang Edition's reflection questions tend toward comprehension-focused prompts. Other scholars have taken the People's Education Press (PEP) high school biology textbook as an example to analyze the important role of the "Thinking and Discussion" section in cultivating students' scientific thinking, including critical thinking and transfer thinking. The study illustrated how 44 instances of the “Thinking and Discussion” section promote the development of students’ thinking skills such as analysis and synthesis, deduction and reasoning, and induction and generalization (Xie et al., 2020). This finding suggests that the design quality of reflective prompts directly affects the effectiveness of cultivating students’ scientific thinking abilities.

Fourth, the two textbooks share common structural characteristics. Spearman correlation analysis revealed strong coherence among evidence, reasoning, and claims, with B1 (completeness of reasoning processes) strongly correlated with both A1 (diversity of evidence sources, r=0.761) and C1 (precision of claim expression, r= 0.632). Ordinal logistic regression further identified B1 as the only significant predictor of claim precision, indicating that the completeness of reasoning chains—not merely the quantity or diversity of evidence—is critical for helping students develop nuanced, conditional understandings of scientific claims.

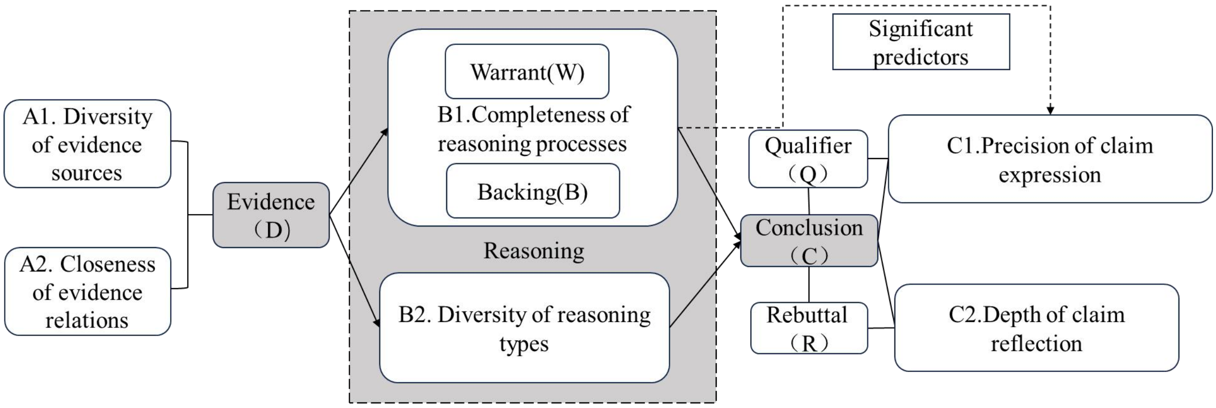

Figure 3. An Evidence-Based Reasoning Literacy Model Grounded in Toulmin's Six Elements

Synthesizing the above findings, this study constructs a relationship model for cultivating evidence-based reasoning (EBR) competence, as shown in Figure 3. The model presents the intrinsic logical relationships among the components: the evidence dimension (A1, A2) provides the foundational material for reasoning, with its quality influencing the construction of reasoning chains; the reasoning dimension (B1, B2) represents the core cognitive activity connecting evidence to conclusions, wherein the completeness of the reasoning process (B1) serves as the key factor determining the conditional qualifiers in claim articulation (C1); within the conclusion dimension, the rigor of claim articulation (C1) corresponds to the qualifier (Q) in the Toulmin model, reflecting the understanding of the scope of knowledge applicability, while the depth of conclusion reflection (C2) corresponds to the rebuttal (R), reflecting critical reflection on knowledge. Notably, C2 shows no significant correlation with the quality of preceding reasoning, exhibiting relative independence, which provides flexible design space for textbook development.

Through this model, and drawing on a substantial body of existing educational research on Toulmin’s Argument Pattern (e.g., Wang and Wang, 2023; Yang and Ni, 2019; Zhang, 2023), this study organically integrates the operationalized elements of Toulmin’s Argument Pattern with textbook text analysis indicators. It thus provides a visual framework for understanding the mechanisms underlying the cultivation of evidence-based reasoning competence in textbooks, and offers a theoretical reference for the subsequent compilation and revision of textbooks.

5.2 Implications for Textbook Design

Based on these findings, three recommendations for science textbook development emerge.

First, *make reasoning chains explicit and complete*. The finding that completeness of reasoning processes predicts claim precision suggests that textbooks should move beyond presenting evidence and claims to explicitly articulate warrants—the logical connections explaining why evidence supports a claim. This may include adding explanatory text such as “This evidence supports the claim because…” after presenting experimental results. While warrants are cognitively demanding to articulate, they are essential for demystifying the reasoning process for students. From the perspective of nature of science (NOS) education, textbooks should help students understand the epistemological underpinnings of science—that is, how scientific knowledge is constructed on the basis of evidence and reasoning—through explicit curriculum design (He, 2018). At the level of cognitive tools, making implicit reasoning processes explicit with the aid of visual thinking scaffolds has been proven to be an effective strategy for cultivating students’ scientific reasoning abilities (Li, 2024).

Second, *diversify reasoning types while ensuring balance*. Both textbooks rely heavily on analogical reasoning, which helps make abstract concepts accessible. This is evidenced by a study on the application of analogical reasoning mechanisms in physics teaching, which found that in the domain of electromagnetism, high school physics textbooks extensively employ analogical reasoning to help students grasp the intrinsic meaning of knowledge and enhance their scientific thinking (Chen et al., 2024).However, the low proportion of causal reasoning—particularly in the Cambridge Edition—suggests that textbooks could strengthen students’ understanding of underlying mechanisms. In their study of refutation texts, Kendeou and colleagues found that analogical reasoning primarily activates text-based local inferences, helping to concretize abstract concepts, whereas causal reasoning drives knowledge-based global inferences, guiding students to systematically explain the mechanisms underlying phenomena (Kendeou et al., 2023). Further research by Grotzer et al. indicates that students often hold deterministic assumptions when reasoning about causal relationships in complex systems, and that instructional guidance in causal reasoning is necessary for them to gradually develop probabilistic causal understanding (Grotzer et al., 2017). An analysis of relational reasoning skills by Vered and colleagues revealed that only 14% of the text presents reasoning skills in an explicit manner; the majority appear implicitly or incompletely explicitly, lacking clear mapping to explain the analogical patterns between source and target domains (Vered et al., 2023).Integrating causal reasoning ("Why does this happen?") alongside analogical reasoning ("What is this like?") and interpretive reasoning ("What does this mean?") may foster more flexible and robust scientific thinking.

Third, *design reflection prompts that cultivate critical awareness*. The finding that reflection depth operates relatively independently of evidence and reasoning quality suggests that textbooks have substantial design flexibility in this area. Davis's research indicates that, compared with generic reflective prompts, directed prompts are more effective in stimulating students' integrative construction of scientific knowledge and in promoting their metacognitive development (Davis, 2003). In their analysis of South African textbooks, Upahi and colleagues found that although the textbooks contain a rich variety of prompt types, most prompts are limited to asking students to "explain what" or "discuss your views," lacking deep guidance for students' reasoning processes and critical thinking (Upahi et al., 2022).Textbooks can systematically incorporate reflection prompts that move beyond comprehension ("What did we learn?") to application ("How does this apply to a new situation?") and critical evaluation ("Are there exceptions? What alternative explanations exist?"). Such prompts can help students develop the capacity to evaluate knowledge claims critically, not merely accept them.

5.3 Limitations and Future Directions

Several limitations should be acknowledged. First, the study focuses on a single topic ("microscopic structure of matter") within two textbooks. Future research could extend the analysis to additional topics and a broader range of textbooks to assess generalizability. Second, the sample size for the Cambridge Edition (12 instructional meaning segments) is smaller than for the Zhejiang Edition (30 segments), which may affect statistical power. Future studies could include additional textbooks or expand the scope of analysis. Third, the analytical framework, while grounded in the Toulmin model, represents one approach to operationalizing EBR. Alternative frameworks might yield complementary insights.

Future research could also examine how students actually engage with EBR opportunities presented in textbooks, investigating the relationship between textbook design and student learning outcomes. Longitudinal studies could explore whether textbooks with more complete reasoning chains, diverse reasoning types, and critical reflection prompts are associated with stronger EBR competencies over time.

6. Conclusion

This study contributes a theoretically grounded, empirically tested framework for analyzing evidence-based reasoning in science textbooks. The findings reveal that while Chinese and international textbooks share fundamental similarities in EBR design, international textbooks demonstrate marginal advantages in diversifying reasoning types, making warrants explicit, and designing critical reflection prompts. Critically, the completeness of reasoning chains emerges as the most significant factor influencing the precision of claim expression, suggesting that textbook authors should prioritize making the logical connections between evidence and claims explicit. By incorporating these design principles, science textbooks can more effectively cultivate the evidence-based reasoning competencies essential for scientific literacy in the 21st century.

Data Availability Statement

The dataset generated and analyzed during the current study is available from the corresponding author upon request.

Conflicts of interest

The authors declare no conflict of interest.

References

Chen, Z. H., Song, X. H., & Yao, J. X. (2024). The thinking mechanism of analogical reasoning and its application in physics teaching. *Physics Teacher*, *45(*5), 9-13. https://doi.org/10.3969/j.issn.1002-042X.2024.05.002

Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences*, *12*(1), 91-142. https://doi.org/10.1207/S15327809JLS1201_4
Duncan, R. G., Chinn, C. A., & Barzilai, S. (2018). Grasp of evidence: Problematizing and expanding the next generation science standards' conceptualization of evidence. *Journal of Research in Science Teaching*, *55*(7), 907-937. https://doi.org/10.1002/tea.21468
Fang, X., Liu, Z. W., Shang, Y. Y., Guan, T. T., Wang, N., Lin, Y. C., & Wang, J. Y. (2023). A comparative study of scientific argumentation elements in Chinese and American junior high school biology textbooks. *World Education Information*. https://doi.org/10.3969/j.issn.1672-3937.2023.10.04
Grotzer, T. A., Solis, S. L., Tutwiler, M. S., & Cuzzolino, M. P. (2017). A study of students‘ reasoning about probabilistic causality: Implications for understanding complex systems and for instructional design. *Instructional Science, 45*(1), 25–52. https://doi.org/10.1007/s11251-016-9389-6
Harsch, R. M., & Kendeou, P. (2023). Learning from refutation texts about scientific topics with analogical and causal explanations. *Contemporary Educational Psychology*, *73*, 102172. https://doi.org/10.1016/j.cedpsych.2023.102172
He, S. L. (2018). A three-dimensional analytical framework of the nature of science and its explicit curriculum design. *Journal of Tianjin Normal University (Basic Education Edition)*, *19*(2), 1-6. https://doi.org/10.16826/j.cnki.1009-7228.2018.02.001
Li, H., & Lu, S. (2024). The conceptual shift and compilation characteristics of Zhejiang Education Edition junior secondary science textbooks. *Teacher Education Forum*, *37*(12), 16-18. https://doi.org/10.3969/j.issn.20955995.2024.12.005
Lin, C. (Ed.). (1990). *Chinese middle school teaching encyclopedia: Education volume*. Shenyang Publishing House. http://interlib.library.hb.cn:8008/opac/book/4003258413
Liu, D. H. (2020). Toulmin's "warrant" and argumentative discourse in modern Chinese. *Contemporary Rhetoric*, (3), 86-95.
Li, Y. (2024). An analysis of primary school science teaching strategies based on thinking visualization. *Primary School Teaching Research*, (14), 4-6. https://doi.org/10.3969/j.issn.1006-284X.2024.14.002
Lu, Z. (2011). *A comparative study of the new and old editions of the People's Education Press chemistry textbooks for senior secondary school grade 1*. Southwest University. https://doi.org/10.7666/d.y1881663
Ministry of Education of the People's Republic of China. (2025). *Science curriculum standards for compulsory education* (2022 Edition, 2025 Revision). Beijing Normal University Press. https://hudong.moe.gov.cn/jyb_xxgk/xxgk/neirong/fenlei/kcjc/kcjc_js/jcjs_kcbz/
Mo, X. (2004). *A study of Toulmin's argumentation model*. South China Normal University.
OECD. (2023). *PISA 2025 science framework*. https://pisa-framework.oecd.org/science-2025/can_eng/
Peng, Y., & Ma, Q. (Eds.). (2010). *Dictionary of logic*. Shanghai Lexicographical Publishing House.
Upahi, J. E., Ramnarain, U., & Ajibola, D. S. (2022). Analysis of writing prompts in South African physical sciences textbooks for Grade 12 learners. *School Science and Mathematics*, *122*(8), 390-401. https://doi.org/10.1111/ssm.12556
Vered, A. A., & Popa, N. L. (2023). Explicit relational reasoning skills: An index for fostering thinking in biology textbooks. *Educatia 21*, (24), 4–17. https://doi.org/10.24193/ed21.2023.24.01
Wang, M. (2019). *New Chinese synonym dictionary*. Shaanxi People's Publishing House.
Wang, S. J., & Wang, C. J. (2023). Analysis and reconstruction of the argumentation structure of "free fall motion" in five versions of textbooks. *Middle School Physics*, *41*(19), 33-36. https://doi.org/10.3969/j.issn.1008-4134.2023.19.007
Wu, H. (2003). Toulmin's model of argument: A critical review of misinterpretations in Chinese literature. *Journal of South China Normal University (Social Science Edition)*, (5), 23–27. https://doi.org/10.3969/j.issn.1000-5455.2003.05.005
Xie, J., Xie, L. C., Deng, G. F., & Huang, S. Q. (2020). The role of the "Thinking and Discussion" section in the new textbook in cultivating students' scientific thinking. *Biology Teaching*, *45*(5), 17-19. https://doi.org/10.3969/j.issn.1004-7549.2020.05.006
Xu, X., & Zhang, Z. (2023). A study of argumentation structures in physics textbooks based on the Toulmin model. *Middle School Physics*, *41*(19), 21–26. https://doi.org/10.3969/j.issn.1008-4134.2023.19.005
Yan, Z. (2019). *Investigation on the current status of "evidence-based reasoning" competency in chemistry*. Shanghai Normal University. http://dx.chinadoi.cn/10.7666/d.Y3543812
Yang, Y. Q., & Ni, J. (2019). Evidence reasoning and model cognition: Connotation analysis and practical strategies. *Chemistry Education*, *40*(23), 23-29. https://doi.org/10.13884/j.1003-3807hxjy.2018080051
Zhang, S. Q. (2021). Decomposing curriculum standards from the perspective of core competencies: Value, elements, and strategies. *Chemistry Teaching*, (2), 7–10. https://doi.org/10.3969/j.issn.1005-6629.2021.02.002
Zhang, S. C. (2012). Cultivation of evidence awareness in inquiry teaching. *Physics Bulletin*, (4), 13-16. https://doi.org/10.3969/j.issn.0509-4038.2012.04.006